

Conference Abstract

# Generating Masks for Image Segmentation in Digitized Herbarium Specimens

Alexander E White<sup>‡</sup>, Rebecca B Dikow<sup>‡</sup>, Makinnon Baugh<sup>§</sup>, Abby Jenkins<sup>§</sup>, Paul B Frandsen<sup>§</sup>

<sup>‡</sup> Data Science Lab, Smithsonian Institution, Washington DC, United States of America

<sup>§</sup> Brigham Young University, Provo UT, United States of America

Corresponding author: Alexander E White ([whiteae@si.edu](mailto:whiteae@si.edu))

Received: 18 Jun 2019 | Published: 26 Jun 2019

Citation: White A, Dikow R, Baugh M, Jenkins A, Frandsen P (2019) Generating Masks for Image Segmentation in Digitized Herbarium Specimens. Biodiversity Information Science and Standards 3: e37479.

<https://doi.org/10.3897/biss.3.37479>

## Abstract

Digitized herbarium images contain complex information unrelated to the shape and color of the specimens represented within them. This information can contribute a substantial amount of noise if one is to use the image as a proxy for pattern, shape, or color of the specimen. Image segmentation, whereby the specimen material is partitioned from the background (e.g., herbarium sheet, label, color ramp), offers one possible solution, yet training data for image segmentation of herbarium specimens is nonexistent. We present a pipeline for generating training data for image segmentation tasks along with a novel dataset of highly resolved image masks segmenting plant material from background noise. This dataset can be used to train neural networks to segment plant material in herbarium sheets more generally, and our method is applicable to other museum data sources where masking may be useful for quantitative analysis of patterns and shapes

## Keywords

digitized herbarium specimens, segmentation, machine learning, image masks

**Presenting author**

Paul B. Frandsen

**Presented at**

Biodiversity\_Next 2019